

Research Article

Image Quality Assessment Based on Joint Quality-Aware Representation Construction in Multiple Domains

Xiaobao Shang,¹ Xinyu Zhao,¹ and Yong Ding² 

¹College of Electrical Engineering, Zhejiang University, Hangzhou 310027, China

²College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China

Correspondence should be addressed to Yong Ding; dingy@vlsi.zju.edu.cn

Received 27 October 2017; Revised 29 January 2018; Accepted 1 February 2018; Published 6 March 2018

Academic Editor: Shang-Hong Lai

Copyright © 2018 Xiaobao Shang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Image quality assessment that aims to evaluate the image quality automatically by a computational model plays a significant role in image processing systems. To meet the need of accuracy and effectiveness, in the proposed method, complementary features including histogram of oriented gradient, edge information, and color information are employed for joint representation of the image quality. Afterwards, the dissimilarities of the extracted features between the distorted and reference images are quantified. Finally, support vector regression is used for distortion indices fusion and objective quality mapping. Experimental results validate that the proposed method outperforms the state-of-the-art methods in terms of consistency with subjective perception and robustness across various databases and different distortion types.

1. Introduction

With the rapid development of multimedia technology, millions of digital images require to be processed; however, any operations on images, such as image acquisition, reproduction, compression, storage, transmission, and restoration, may introduce noise into images [1, 2]. Image quality assessment (IQA) is employed to estimate the degree of distortion. That is, IQA plays a pivotal part in evaluating or monitoring the performance of an image processing system. Visual perception is the result of interaction between human being and environment, and it is difficult to define image quality exactly. The most accurate estimation method for image quality is subjective IQA carried out by human observers. However, subjective quality methods are costly, time-consuming, and impractical as they cannot be integrated within real-world systems for real-time visual quality monitoring and controlling [3, 4]. It triggers the need to develop reliable objective IQA methods that are consistent with subjective human evaluation.

Traditional objective IQA methods like Peak Signal-to-Noise Ratio (PSNR) and Mean Squared Error (MSE) [2] are simple and with clear physical meanings, but they are

blamed for their low accuracy because they just measure the statistical information in images. Thereby, in recent decades, an increasing number of researchers make efforts to develop effective and efficient methods to evaluate the image quality automatically. Ideally, perceptual quality is obtained by mimicking the image perception mechanism of human visual system (HVS); nevertheless, due to the complexity and limited understanding of HVS, it is almost impossible to completely imitate HVS [5]. State-of-the-art methods turn to the other direction that manages to capture the statistical properties (features) which represent the information that HVS is interested in and are closely relevant to the image inherent quality and map them to the perceptual quality. Therefore, it is a significant problem for us to extract the effective image feature. Furthermore, in most cases, a single measurement cannot provide sufficient information for quality prediction. It is proved that exploiting the image information jointly in image space, scale, and orientation domains can provide richer clues, which are not evident in any single domain. That is to say, quality-related representation should be obtained by considering joint occurrences of two or more features [6]. Saha et al. put forward a method to evaluate the image quality based on extracting multiple features which are composed

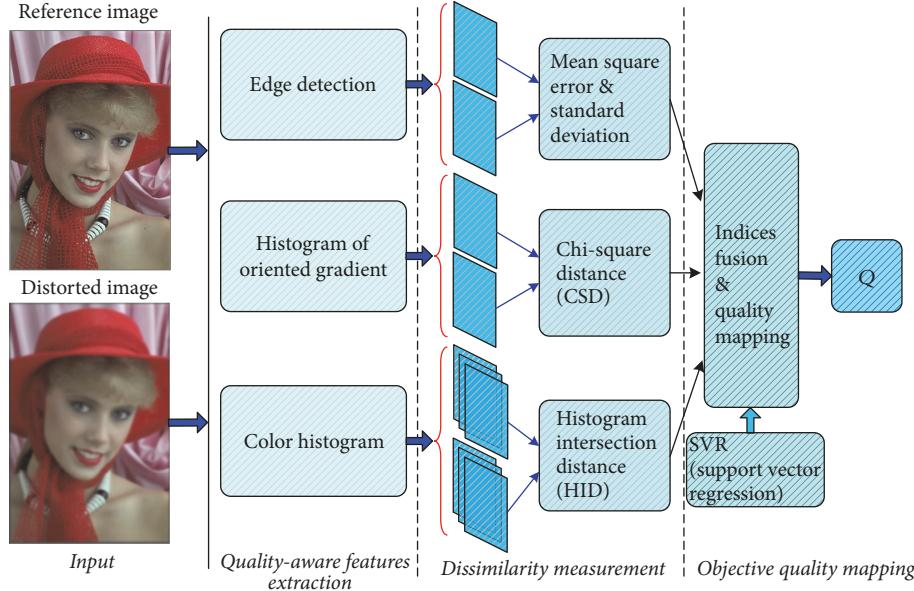


FIGURE 1: Framework of the proposed method.

of gradient difference, contrast difference, and saliency map difference [7]. And, in [6], an approach of complementary representation is proposed by employing log-Gabor filters to model the perceptual characteristics and local derivative pattern to describe detailed texture information.

Furthermore, an effective IQA method is extremely desirable to be (i) consistent with subjective sensation of HVS and (ii) stable across different types of distortions. To tackle the challenge mentioned above, a method based on complementary feature extraction and fusion is proposed in this paper. The framework of the proposed method is shown in Figure 1. Firstly, quality-aware features on three domains are extracted to construct the complementary representation. Secondly, the difference of each feature between the reference and distorted images is quantified. Such differences that reflect the degree of the quality degradation are taken as the distortion indices. Thirdly, all the distortion indices are mapped into an objective quality score, where Support Vector Machine (SVM) is adopted to learn a regression model.

The main contributions of this paper lie in the following: a complementary representation of image quality is proposed based on joint quality-aware feature extraction, where multiscale histograms of oriented gradient (HOG) are used for local structure information description, multiscale edge information is employed to describe the global information, and, meanwhile, color histograms are used to describe the color information. With the complementary quality-aware feature extraction and SVM-based fusion, a high-efficiency IQA method which performs well in terms of consistency with subjective assessment and robustness across different databases and distortions types is put forward.

The organization of the remainder of this paper is as follows: In Section 2, a brief review of IQA and related work is given. Section 3 presents the multiscale quality-aware feature extraction and the corresponding dissimilarity quantification. Section 4 introduces the feature fusion and quality

mapping strategy. Experimental results and discussions are presented in Section 5, and, finally, in Section 6, we draw the conclusion of this paper and indicate the future work.

2. Related Work

Generally, the intrinsic principle of IQA is to measure the degree of perceptual quality degradation by assessing the difference or dissimilarity between the distorted image and its corresponding reference image [3]. The scheme of a generic IQA method is shown in Figure 2, where there are three stages. In the first stage, features that can reflect the image quality are extracted by different algorithms. Then, difference or dissimilarity of each feature between the distorted image and its corresponding reference is quantified in the second stage. Such differences or dissimilarities are regarded as the distortion indices to measure the degradation of image quality. Finally, in the third stage, all the distortion indices are fused together and mapped into an objective quality score.

Feature extraction is quite important in image quality assessment. Over the last decades, various effective feature extraction methods have been explored in literature, which are conducted in two directions [8]. Since images are ultimately viewed by human beings, by taking advantage of the limited understanding of human visual system (HVS), some meaningful HVS-oriented methods capture the features by simulating the way that human beings perceive the image [9]. Study on neural science has discovered that the primary and secondary visual cortices (V1 and V2) are the earliest receptors for signal and play the most important part in generating vision [10]. Recent HVS-based methods try to incorporate just noticeable difference (JND) [11], contrast sensitivity function (CSF) [12], contrast masking [13], temporal masking, visual attention, and saliency map into the perceptual feature extraction. For example, the well-known method is visual information fidelity (VIF) [14] which incorporates the HVS

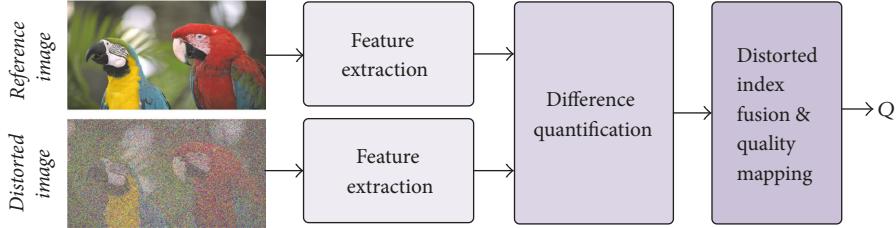


FIGURE 2: Scheme of an IQA method.

model into information fidelity criterion (IFC) [15]. Since images are naturally multiscale in HVS perception, some weighted multiscale methods at multiple resolutions cohering with human perception are explored [16, 17]. Furthermore, it has been found that Gabor filters have excellent properties and the shapes of Gabor wavelets are similar to the receptive fields of simple cells in the primary visual cortex. Thus, visual features are extracted by 2D Gabor filter to reflect the nonlinear mechanism of HVS [18]. Instead of simulating the functional components of the low-level HVS, some high-level aspects of HVS, such as visual attention and visual saliency, are considered into feature extraction [9, 19]. However, it is very difficult to model the complex and rigorous HVS well relying on the limited understanding upon it. Besides, the high computational complexity also limits the application of these methods.

In this scenario, most of state-of-the-art methods turn to the other direction that attempt to extract the statistical properties (features) of an image that are closely related to the image inherent quality [6]. Such methods are inclined to effectively extract quality-aware features and have achieved notable success. Based on the hypothesis that HVS is highly adapted for extracting structural information from images, structural similarity (SSIM) index is quite attractive, in which the luminance, contrast, and structure information are involved for perceptual quality estimation [20]. Furthermore, several extensive methods have been proposed to improve the performance [17, 21–23]. Recent studies prove that exploiting the image information jointly in color space, scale, and orientation domains can provide richer clues, which are not evident in any single domain [5]. Therefore, multidomain feature extraction by considering joint occurrences of two or more features is necessary and constructive [6, 10].

In the respect of image quality assessment, if there are some distortions introduced in an image, they may inevitably affect the statistic characteristics of features. Intuitively, the features' statistic characteristics of the distorted image are quite different from that of the reference image. Quantifying deviations from the regularity appropriately is validated as a useful way for assessing the perceptual quality. Thus, in the stage of difference quantification, the dissimilarity of each feature between the reference and distorted images is computed. There are many common approaches to quantify the difference between two features, such as Euclidean distance, histogram intersection distance (HID), histogram quadratic (cross) distance, Chi-square, and correlation coefficient [24, 25]. In [6], correlation coefficient is employed to quantify the

difference between the gradient magnitude and orientation maps, chi-square distance is used to quantify the difference between the energy maps, and HID is adopted to quantify the difference between the local pattern maps.

As mentioned above, the dissimilarities of features between the distorted and reference images can reflect the image quality degradation. It is reasonable to take them as distortion indices. That is, the quantification can be regarded as a good approximation of perceived distortion in image quality. In the stage of indices fusion and quality mapping, all the distortion indices should be fused together and mapped into an objective score. Average fusion which calculates the mean value as the overall quality score is simple and widely used, but its accuracy is generally blamed. Weighted-based fusion strategy is another widely used approach because various features may contribute differently to the final quality. More recently, researchers tend to construct more complex weighting function or employ machine learning techniques to develop the fusion scheme [4]. K-nearest neighbor (KNN), convolutional neural network (CNN), and SVR are the most commonly used machine learning tools [26].

3. Quality-Aware Feature Extraction and Dissimilarity Quantification

It is claimed that human eyes are sensitive to less complex features which are localized, oriented, and bandpass and interested in color and structure information [27]. In this paper, we attempt to capture complementary quality-aware features from different domains. Quality-aware features represent the characteristics which are closely related to the quality of image. In addition, quality-aware features are sensitive to the degradation of image quality, and the changes of these features due to the image distortions are consistent with the human eye vision quality perception.

3.1. Orientation Feature Extraction and Dissimilarity Quantification. With regard to feature extraction, local description has received a lot of attention in recent years which performs very well in many image applications including image retrieval, object recognition, and texture analysis [6]. Such success inspires us to introduce local features extracted by HOG into image quality assessment.

As shown in Figure 3, there are five steps for calculating HOG of an image. In the first step, a pair of masks $[-1, 0, 1]$ and $[-1, 0, 1]^T$ are used to obtain the gradient magnitude and orientation maps because the weights of positions in

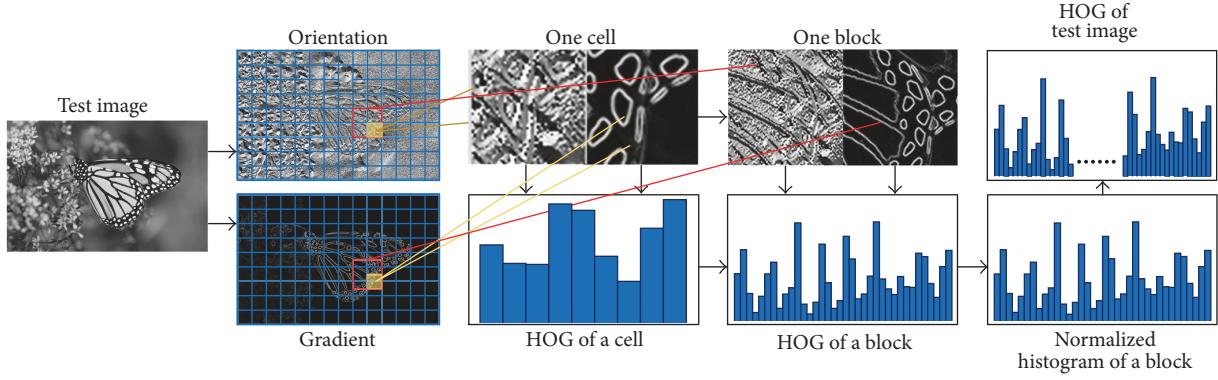


FIGURE 3: Illustration of HOG process.

the masks are in connection with their relevant distance to center pixel and thereby can achieve relatively better results. Image segmentation is the second step, in which the image is divided into a fixed number of 8×8 small connected regions called cells. Histogram accumulation is the third step, where each cell is discretized into several angular bins according to the gradient orientation, and a local HOG over all pixels within one cell is accumulated [28]. The weighted votes of the local HOG are generated by the gradient magnitudes. In this paper, the range of orientations is from 0° to 180° , and the gradient orientation is divided into nine bins. The next step is grouping histograms of cells into blocks, in which the histograms of adjacent cells (2×2 cells) are concatenated to constitute a descriptor of one block. After grouping, the values of histograms within the blocks are normalized in order to avoid contrast bias. The final step is concentrating total histograms of blocks into a combined vector which is regarded as the final feature descriptor.

The degree of corruption can be regarded as the approximate value of the image quality degradation because an image with better quality has smaller divergence with its perfect image. Chi-square distance is adopted as the dissimilarity measure to construct the first kind of distortion index.

$$D_{\text{HOG}} = \frac{1}{M} \sum_{m=1}^M \frac{(\text{HOG}_R(m) - \text{HOG}_D(m))^2}{\text{HOG}_R(m) + \text{HOG}_D(m)}, \quad (1)$$

where M is the total number of HOG bins within an image and $\text{HOG}_R(m)$ and $\text{HOG}_D(m)$ denote the values of reference and distorted images in the same HOG bin, respectively.

3.2. Edge Feature Extraction and Dissimilarity Quantification. It is well known that human eye is significantly sensitive to edge and contour information of an image; therefore, edge information is very important for fully exploiting the potential inherent quality. There are many useful tools for edge information detection, such as Canny, Sobel, Scharr, and Prewitt operators. Prewitt operator is selected to obtain the edge information in this paper. The Prewitt operators

in horizontal (x) and vertical (y) directions are shown as follows:

$$P_x = \begin{bmatrix} \frac{1}{3} & 0 & -\frac{1}{3} \\ \frac{1}{3} & 0 & -\frac{1}{3} \\ \frac{1}{3} & 0 & -\frac{1}{3} \end{bmatrix}, \quad (2)$$

$$P_y = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 0 \\ -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} \end{bmatrix}.$$

An edge map is produced by the following:

$$E_d = \sqrt{(D \otimes P_x)^2 + (D \otimes P_y)^2} \quad (3)$$

$$E_r = \sqrt{(R \otimes P_x)^2 + (R \otimes P_y)^2},$$

where symbol “ \otimes ” denotes a convolution operator and E_r and E_d represent the edge maps of the distorted image and its reference image, respectively. As an example, a reference image named “parrots” taken from LIVE database [29] and its corresponding detected edge information are shown in Figures 4(a)–4(e). For comparison, the relevant results of its distorted version with the distortion of white Gaussian noise (WN) are shown in Figures 4(f)–4(j).

Ideally, the edge map should have a continuous outline because the edge distribution of a perfect image is highly organized; in comparison, perfection of the outline for a distorted image is destroyed [8]. From Figure 4, it can be apparently observed that the outline in Figure 4(e) is much clearer than that in Figure 4(j). Moreover, the degree of degradation reflects the distortion level of the images. Edge similarity map between the distorted and reference images is calculated in a pixel-wise manner, which is defined as follows:

$$\text{ES}(i) = \frac{2E_r(i)E_d(i) + c}{E_r(i)^2 + E_d(i)^2 + c}, \quad (4)$$

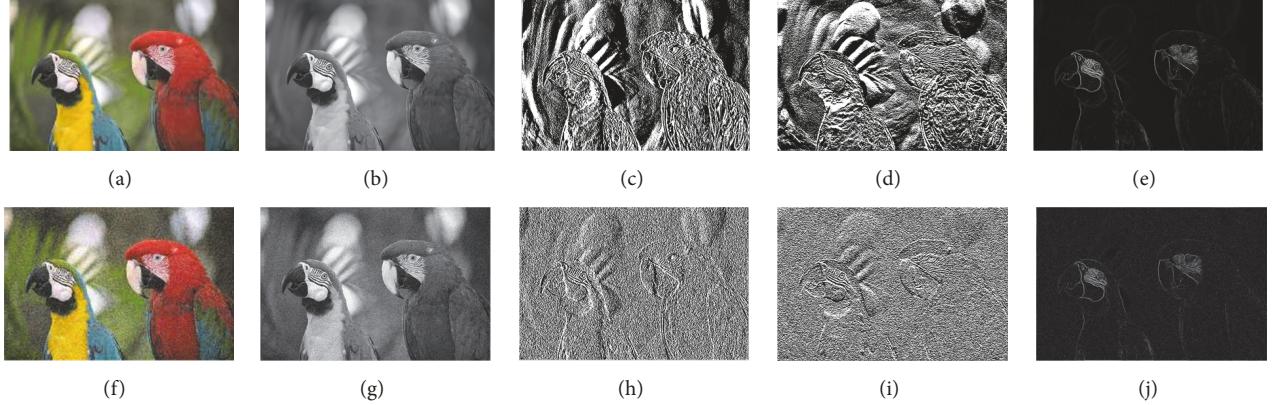


FIGURE 4: (a) An original reference image, (b) gray image, (c) horizontal image with Prewitt operator, (d) vertical image with Prewitt operator, and (e) edge map of image. ((f)–(j)) A WN distorted version of the image and its corresponding results.

where c is a positive constant utilized for numerical stability and $\text{ES}(i)$ denotes the dissimilarity of i th pixel between reference and distorted images. In this paper, we take the mean value EA and standard deviation ED of ES to act as two indices of image quality:

$$\begin{aligned} \text{EA} &= \frac{1}{N} \sum_{i=1}^N \text{ES}(i) \\ \text{ED} &= \sqrt{\frac{1}{N} \sum_{i=1}^N (\text{ES}(i) - \text{EA})^2}, \end{aligned} \quad (5)$$

where N is the total number of pixels contained in an image.

3.3. Color Feature Extraction and Dissimilarity Quantification. HOG and edge detection just extract the features in gray channel. However, color plays a crucial role in the perception of an image [5]. To extract the quality-aware feature sufficiently, in this paper, color histograms based on HSV space are employed to describe color information. Before the color information extraction, images are converted from RGB space into HSV space:

$$\begin{aligned} H &= \cos^{-1} \left\{ \frac{(1/2)[(R-G)+(R-B)]}{\sqrt{(R-G)^2 + (R-B)(G-B)}} \right\} \\ S &= 1 - \frac{3}{R+G+B} [\min(R, G, B)] \\ V &= \frac{1}{3} (R+G+B), \end{aligned} \quad (6)$$

where H , S , and V are hue, saturation, and the value of intensity. In general, H , V , and S are discretized into 8, 3, and 3 bins according to their magnitude, respectively. After discretization, H is usually calculated an integer between 0 and 8, V is usually calculated an integer between 0 and 2, and S is usually calculated an integer between 0 and 2.

Given an image I , the color histogram (HI) that describes the global distribution of pixels within an image can be represented as follows:

$$\text{HI} = \{h_1, h_2, h_3, \dots, h_N\}, \quad (7)$$

where N is the number of bins that color histograms are categorized into. In HSV space, h_i is computed as

$$\begin{aligned} h_i &= \sum_{x=0}^K \sum_{y=0}^T \text{sgn}(A(x, y)) \\ \text{sgn}(A(x, y)) &= \begin{cases} 1, & A(x, y) = i - 1 \\ 0, & \text{otherwise}, \end{cases} \end{aligned} \quad (8)$$

where $A(x, y) = \epsilon \times H(x, y) + \beta \times S(x, y) + \chi \times V(x, y)$, K, T denote the horizontal and vertical pixel numbers contained in an image, and ϵ, β , and χ are the parameters should be defined. In this paper, $\epsilon = 16$, $\beta = 4$, and $\chi = 1$.

The dissimilarity of color histograms between the distorted and reference images is then quantified by histogram intersection distance. Such dissimilarity is regarded as the third distortion index.

$$D_{\text{COL}} = \frac{\sum_{n=1}^N \min(\text{HI}_R(n), \text{HI}_D(n))}{\sum_{n=1}^N \text{HI}_R(n)}, \quad (9)$$

where HI_D and HI_R represent the color histograms of the distorted image and its reference, respectively, and N is the number of bins.

4. Multiscale Features Fusion and Objective Quality Mapping

Multiscale approach is an advantageous way to incorporate image details at different resolutions. It is discovered that human eyes can easily identify and process the natural images at different scales [1]. Therefore, processing a natural image at different scales can increase the flexibility and adaptation for the image quality evaluation [30]. In our IQA model, three

types of features mentioned above in multiscale are extracted from the original scale image and the corresponding down-samples.

With all these quantified multiscale distortion indices, many methods can be employed to construct a function to synthesize and map them into objective quality scores. Traditionally, linear approaches are usually used to fuse and map the indices; for example, use weighted average values of indices [31]. In recent years, machine learning techniques are introduced to derive the single metric, such as KNN, deep learning, and SVR. The theory of KNN is simple, but its accuracy is unsatisfactory, especially for large-scale data. Deep learning is more precise, but it spends a lot of time building topological structure and determining the values of parameters. In addition, an efficient network model is based on a large amount of training data. SVR is an accurate tool for constructing the regression function in IQA [8]. Under the condition of small samples, it is difficult to obtain ideal results because deep learning usually generates overfitting problem. Since the experimental data provided by the databases is limited, SVR is employed to construct a function to synthesize and map them into objective quality score. To be specific, ε -SVR with the radius-basis function (RBF) kernel is employed in the proposed method [8, 32].

$$Q = F_{\text{SVR}} \{D_{\text{HOG}}, \text{EA}, \text{ED}, D_{\text{COL}}, D_{\text{HOG}2}, \text{EA}_2, \text{ED}_2, \\ D_{\text{COL}2}\}, \quad (10)$$

where $D_{\text{HOG}2}$, EA_2 , ED_2 , and $D_{\text{COL}2}$ represent the indices of orientation, edge, and color information in the second scale, respectively. The framework of the dissimilarity quantification and quality mapping is shown in Figure 5, where L is the original scale image and $2\downarrow$ represents its reduced scale image with downsampling by 2, “HOG” denotes the HOG, “GM” is the edge map, “COL” means the color histogram, and suffixes “_ref” and “_dis” imply the features are extracted from a reference image or a distorted image.

ε -SVR is used as the machine learning tool to find the best F_{SVR} in (11) which minimizes the dissimilarity between subjective and objective quality scores. Function F can be defined as

$$F(x) = \omega^T \varphi(x) + \delta, \quad (11)$$

where ω and δ are weight vector and deviation factor, respectively, which are introduced to insure the dissimilarity between subjective score S and objective score Q less than ε for all training data. $\varphi(\bullet)$ represents a nonlinear function which maps x into a nonlinear space, and ω is represented as

$$\omega = \sum_{i=1}^{n_{\text{SVR}}} (\eta_i^* - \eta_i) \varphi(x_i), \quad (12)$$

where η_i^* and η_i ($0 \leq \eta_i^*, \eta_i \leq C$, and C is the parameter of tradeoff error) represent the Lagrange multipliers and n_{SVR}

denotes the amount of support vectors [33]. Function F can be rewritten as

$$F(x) = \sum_{i=1}^{n_{\text{SVR}}} (\eta_i^* - \eta_i) \varphi(x_i) \varphi(x) + \delta \\ = \sum_{i=1}^{n_{\text{SVR}}} (\eta_i^* - \eta_i) \phi(x_i, x) + \delta, \quad (13)$$

where $\phi(x_i, x)$ is the radius-basis kernel function:

$$\phi(x_i, x) = e^{-\alpha \|x_i - x\|^2}, \quad (14)$$

where α is the width of the kernel function. A high value of α will lead to overfitting problem, which reduces the generalization ability of the IQA model. On the contrary, the accuracy of the IQA model will be degraded if the value of α is too low. Therefore, the value of α has influence on quality assessment.

5. Experimental Results and Discussion

5.1. Experimental Setup. Experiments are conducted on four large-scale image databases, LIVE [29], TID2008 [34], TID2013 [35], and CSIQ [36]. Hundreds even thousands of distorted images which are corrupted by various types of distortions are contained in each database. And subjective ratings in each database which are in the form of either mean opinion score (MOS) or difference of mean opinion score (DMOS) give the perceptual quality for each distorted image. Detailed information of the databases is shown in Table 1.

For performance evaluation, Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-order Correlation Coefficient (SRCC), and Root Mean Square Error (RMSE) between objective and subjective quality scores are adopted as the evaluation metrics, where PLCC and RMSE are used to assess prediction accuracy, and SRCC is adopted to evaluate prediction monotonicity. It deserves to be mentioned that, before computing PLCC and RMSE, a nonlinear mapping is carried out between subjective and objective scores using the modified logistic regression model [20] because different databases adopt different schemes to quantify subjective IQA results. The nonlinear mapping function is defined by

$$Q_P(Q) = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(Q - \beta_3))} \right) + \beta_4 Q \\ + \beta_5, \quad (15)$$

where Q indicates the raw objective score and β are the parameters to describe the relationship between Q and DMOS. The associated quality difference scores of distorted images are available for fitting the five parameters and establishing the prediction of the nonlinear mapping function. Higher PLCC and SRCC values yet lower RMSE mean better performance of the IQA method.

It should be noted that fivefold cross-validation is adopted. Images from one database are split randomly into two no-overlapped subsets: 80% used for training and 20%

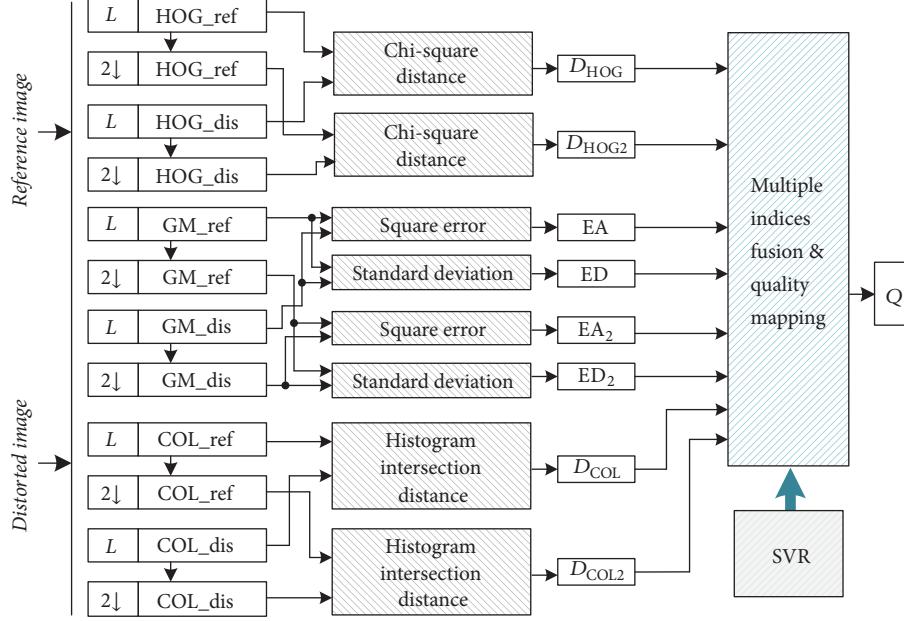


FIGURE 5: Structure diagram of dissimilarity quantification and quality mapping.

TABLE 1: Fundamental information about the four largest databases used in our experiments.

Database	Number of reference images	Number of distorted images	Distorted types	Resolution
LIVE	29	982	5	Various
TID2008	25	1700	17	512 × 384
TID2013	25	3000	24	512 × 384
CSIQ	30	866	6	512 × 512

used for testing. A predicted quality score is acquired for each image after testing, and then the performance is assessed based on all quality scores. In order to eliminate the performance bias, random-test is repeated 1000 times and the mean value across these 1000 iterations is reported as the final result in this paper.

5.2. Overall Performance across Databases. Table 2 shows the overall performance comparison with state-of-the-art IQA methods including MCSD [1], IFS [4], LCSIM3 [5], CLFE [6], CLR [8], VSI [9], SSIM [20], GMSD [25], ESIM2 [37], and QASD [38] on different databases. For SSIM, an alternative framework based on structural similarity is introduced to evaluate the quality of images. With respect to GMSD, gradient magnitude similarity deviation is presented to construct the IQA model. Numbers of new methods are developed to extract quality-aware features to assess the image quality recently; for example, MCSD is utilized to represent the image quality. In addition, an adaptive subdictionaries index is put forward for IQA in [38]. LCSIM3 [5], IFS [4], CLFE [6], CLR [8], ESIM2 [37], and the proposed method try to improve the accuracy and robustness across different databases by multiple feature extraction, indicating that complementary features construction is a promising solution for effective IQA development. However, one of the

vital problems for multiple features based IQA methods is to find proper fusion function. With the machine learning tool SVR, our method achieves superior performance. In Table 2, the best results on each database are highlighted in boldface. It is obvious that the performance of the proposed method is superior to all the other compared IQA methods on TID2008 and TID2013 and near to the best on LIVE and CSIQ, while, for other competitors, they may perform well on one or two databases but poor on other databases [39]. For example, earlier IQA methods, like SSIM, perform well on only one or two databases, and CLFE [6] works well on LIVE, but it performs poorly on TID2008 and TID2013. The earlier proposed IQA methods perform poorly because only one feature is employed to describe the degradation of image quality in the earlier IQA methods. However, in recent years, many IQA methods based on combinational and joint representation of image quality are proposed. For CLR and CLFE, edge feature, local spatial-frequency feature, and texture feature are used to complementarily represent the image quality, and SVR is employed to fuse the multiple distortion indices. In respect of LCSIM3 and ESIM2, several earlier IQA approaches are used to jointly evaluate the image quality. From Table 2, it can be seen that the methods that employ complementary features can obtain the outcome expected. There are only five distortion types contained in the

TABLE 2: Performance comparison on different databases.

	SSIM	IFS	ESIM2	CLR	QASD	LCSIM3	CLFE	VSI	GMSD	MCSD	Ours
LIVE											
PLCC	0.9449	0.9613	0.9613	0.9775	0.9602	0.9722	0.9829	0.9482	0.9603	0.9675	0.9801
SRCC	0.9479	0.9645	0.9598	0.9661	0.9646	0.9662	0.9755	0.9524	0.9603	0.9668	0.9749
RMSE	8.9455	7.5296	7.6656	5.0492	7.6334	7.0457	4.1701	8.6816	7.6214	6.9079	4.5550
TID2008											
PLCC	0.7732	0.8762	0.9026	0.8885	0.8877	0.9107	0.8881	0.8762	0.8788	0.8844	0.9283
SRCC	0.7749	0.8840	0.8964	0.8806	0.8899	0.9113	0.8880	0.8979	0.8907	0.8911	0.9245
RMSE	0.8511	0.6468	0.5947	0.6106	0.6178	0.5525	0.6120	0.6466	0.6404	0.6263	0.4966
TID2013											
PLCC	0.7895	0.8769	0.8804	0.8817	0.8894	0.8307	0.8939	0.9000	0.8553	0.8565	0.9269
SRCC	0.7417	0.8510	0.8990	0.8516	0.8557	0.8659	0.8802	0.8965	0.8044	0.8089	0.9114
RMSE	0.7608	0.5959	0.5430	0.5862	0.5667	0.6201	0.6552	0.5404	0.6423	0.6399	0.4641
CSIQ											
PLCC	0.8613	0.9192	0.9620	0.9606	0.9466	0.9624	0.9680	0.9279	0.9541	0.9560	0.9689
SRCC	0.8756	0.9310	0.9606	0.9573	0.9516	0.9704	0.9627	0.9423	0.9570	0.9592	0.9639
RMSE	0.1334	0.1034	0.0729	0.0728	0.0846	0.0634	0.0650	0.0979	0.0786	0.0770	0.0646
w.a.											
PLCC	0.8181	0.8950	0.9091	0.9083	0.9071	0.8909	0.9155	0.9047	0.8902	0.8935	0.9408
SRCC	0.7990	0.8872	0.9156	0.8903	0.9047	0.9065	0.9074	0.9113	0.8704	0.8738	0.9362
RMSE	1.9287	1.5838	1.5624	1.1939	1.5311	1.4926	1.0931	1.7304	1.6139	1.5019	1.0332

TABLE 3: Statistical significance tests.

	SSIM	IFS	QASD	CLR	CLFE	LCSIM3	ESIM2	VSI	GMSD	MCSD
LIVE	1	1	1	0	0	0	1	1	1	1
TID2008	1	1	1	1	1	1	1	1	1	1
TID2013	1	1	1	1	1	1	1	1	1	1
CSIQ	1	1	1	0	0	1	1	1	1	1

TABLE 4: SRCC of the proposed method with others for all five individual distortion types on LIVE database.

	SSIM	SURF	GLD-SR	CLR	BIFS	LCSIM	CSV	VSI	GMSD	MCSD	Ours
JPEG	0.9614	0.9760	0.9827	0.9784	0.9810	0.9818	0.9600	0.9604	0.9711	0.9720	0.9808
JP2K	0.9764	0.9560	0.9730	0.9768	0.9840	0.9721	0.9800	0.9761	0.9782	0.9790	0.9773
WN	0.9694	0.9800	0.9861	0.9777	0.9910	0.9873	0.9910	0.9835	0.9737	0.9809	0.9895
GB	0.9517	0.9510	0.9499	0.9737	0.9550	0.9589	0.9730	0.9527	0.9567	0.9528	0.9847
FF	0.9556	0.9510	0.9541	0.9662	0.9320	0.9694	0.9500	0.9430	0.9416	0.9527	0.9817
w.a.	0.9636	0.9631	0.9701	0.9749	0.9702	0.9742	0.9708	0.9638	0.9655	0.9684	0.9824
σ	0.0101	0.0141	0.0165	0.0050	0.0246	0.0111	0.0162	0.0156	0.0150	0.0112	0.0027

LIVE database. Therefore, numbers of IQA methods perform well in the LIVE database, especially CLFE. It means that edge and texture features are enough to describe the degradation of the image quality in the LIVE database. However, other features are acquired to be extracted to represent the image quality because edge and texture features will not be able to accurately represent all type of distortion. Thus, color histograms are employed to describe color information in this paper. In Table 2, it is obvious that our algorithm has strong robustness. CLFE only performs well in the LIVE database. Furthermore, the weighted averages (w.a.) of PLCC, SRCC, and RMSE are used to show experimental results intuitively, where the number of images in each database is taken as the weight. Clearly, the proposed method achieves the highest average PLCC (0.9408) and SRCC (0.9362) and lowest RMSE (1.0332) outperforming all the listed methods.

In addition, to analyze the statistical significance of the proposed method, a left-tail F -test [40] with significant level at 1% is carried out. It should be mentioned that the prediction residuals of the objective scores after nonlinear mapping are used by F -test. Table 3 lists the results, where the symbols “1”; “0”; or “-1” suggest that our method is statistically better, indistinguishable, or worse than the other methods. It can be found that our method is significantly better than other methods on TID2008 and TID2013 databases and has a small advantage on LIVE and CSIQ. There are only several indicators that appear as 0: comparable to CLR and CLFE on LIVE and CSIQ, comparable to LCSIM3 on LIVE. However, CLR and CLFE perform so poorly on TID2008 and TID2013 databases. Consequently, the proposed method has performance superiority over all listed methods because most of the values in Table 3 appear as 1.

Besides, a good IQA method should predict the image quality consistently across repeating many times. Thus, the box-plot of PLCC generated during the cross-validation period on different databases is presented in Figure 6 to demonstrate the stabilization of the proposed method. In

Figure 6, for one box, the center mark is median, the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually. It is obvious that our method is concentrated and has small whiskers in different databases, which implies it achieves a stable and impressive performance.

5.3. Performance on Individual Distortion Type. For verifying the robustness of the proposed method further, the prediction accuracy on individual distortion type is conducted. A method with high robustness should not only attain good accuracy on all distortion types but also achieve stability among different distortion types. LIVE database is used to conduct the experiments, which contains five distortion types, including JPEG2000 (JPEG), JPEG, white Gaussian noise (WN), Gaussian blur (GBLUR), and fast-fading (FF).

Table 4 lists the SRCC comparison of the proposed method with others for all five individual distortion types on LIVE database. The compared methods include MCSD [1], BIFS [3], LCSIM [5], GLD-SR [7], CLR [8], VSI [9], SSIM [20], GMSD [25], CSV [40], and SURF [41]. In Table 4, the top performance for each distortion type is highlighted in boldface, too. It can be found that the proposed method obtains a remarkable high SRCC across five distortion types. The excellent performance across different distortion types lies in the following: (i) WN can be detected by edge features because WN always results in high frequency component decrease which makes the edge information destroyed. It is the same with JP2K and GB, where they cause low frequency component increase and edge being unclear. (ii) For JPEG, since new edge appears in the boundary of the blocks with the edge of original image declining, block-based feature extraction method is fit for JPEG. Therefore, in the proposed method, JPEG can be detected by the HOG operation. Although HOG is mainly used to detect local distortion, it

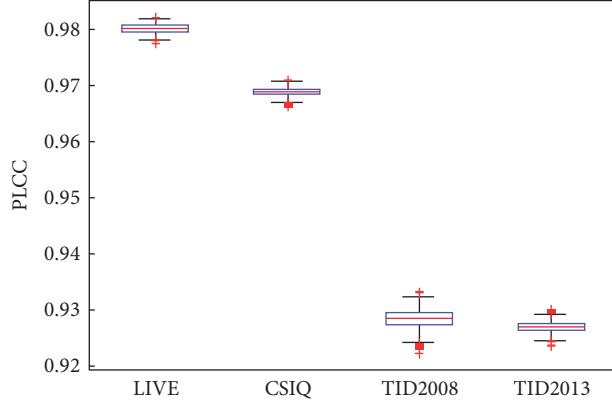


FIGURE 6: Box-plots of cross-validation phase.

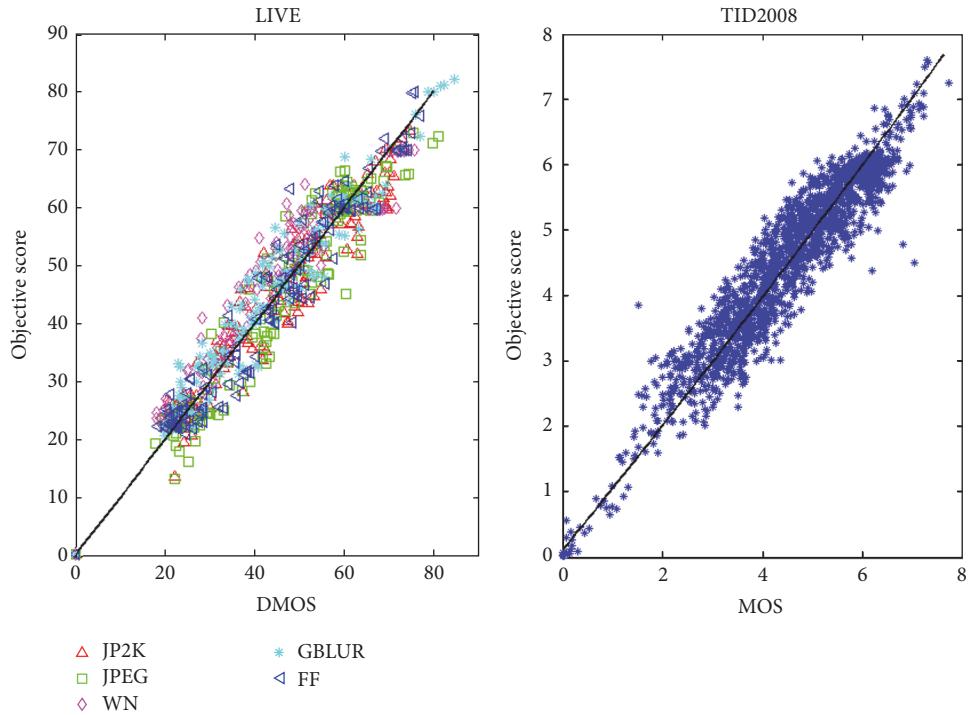


FIGURE 7: Scatter plots on LIVE and TID2008 databases.

is also fit for FF because FF is a local distortion appearing on random area in an image.

The practical image processing system is unpredictable and may encounter complex situations, so it is important for an IQA method to maintain stability. In order to more directly validate the performance across various distortion types, weighted average (w.a.) and standard variation (σ) of SRCC of five distortion types are shown in Table 4. For w.a. calculation, the number of images contained for each distortion type is taken as the weight, similar to what it does in Table 2. High value of w.a. means excellent performance; meanwhile, low value of σ reveals that the method has a power to maintain a stability in evaluating images contaminated by different types of distortions. Obviously, our method has largest w.a. (0.9824) and comparatively small σ (0.0027), which means that the

proposed method can deal with all the listed distortion types very well.

To further prove the performance of our method, the scatter plots on LIVE and TID2008 databases are shown in Figure 7, where the objective score shown in vertical axis is generated by a nonlinear mapping function, horizontal axis is subjective quality scores, and every dot represents an image in databases. Obviously, the objective scores show a large correlation with subjective image quality scores, as the distribution of the scatter plots is closely next to fitted curve.

6. Conclusion

In this paper, a complementary image quality representation is introduced to develop a highly effective IQA method,

which consists of orientation and edge and color information. Conducted on four public databases, the experimental results show that the proposed method achieves an excellent performance in terms of both prediction accuracy and robustness across different databases and distortion types, which provides an insightful and promising solution for high performance IQA development.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

The authors would like to acknowledge the support from the Fundamental Research Funds for the Central Universities.

References

- [1] T. Wang, L. Zhang, H. Jia, B. Li, and H. Shu, "Multiscale contrast similarity deviation: An effective and efficient index for perceptual image quality assessment," *Signal Processing: Image Communication*, vol. 45, pp. 1–9, 2016.
- [2] S. Du, Y. Yan, and Y. Ma, "Blind image quality assessment with the histogram sequences of high-order local derivative patterns," *Digital Signal Processing*, vol. 55, pp. 1–12, 2016.
- [3] F. Gao and J. Yu, "Biologically inspired image quality assessment," *Signal Processing*, vol. 124, pp. 210–219, 2016.
- [4] H.-W. Chang, Q.-W. Zhang, Q.-G. Wu, and Y. Gan, "Perceptual image quality assessment by independent feature detector," *Neurocomputing*, vol. 151, no. 3, pp. 1142–1152, 2015.
- [5] M. Oszust, "Full-reference image quality assessment with linear combination of genetically selected quality measures," *PLoS ONE*, vol. 11, no. 6, article e0158333, 2016.
- [6] Y. Zhao, Y. Ding, and X. Zhao, "Image quality assessment based on complementary local feature extraction and quantification," *IEEE Electronics Letters*, vol. 52, no. 22, pp. 1849–1851, 2016.
- [7] A. Saha and Q. M. J. Wu, "Full-reference image quality assessment by combining global and local distortion measures," *Signal Processing*, vol. 128, pp. 186–197, 2016.
- [8] Y. Ding, Y. Zhao, and X. Zhao, "Image quality assessment based on multi-feature extraction and synthesis with support vector regression," *Signal Processing: Image Communication*, vol. 54, pp. 81–92, 2017.
- [9] L. Zhang, Y. Shen, and H. Li, "VSI: a visual saliency-induced index for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 23, no. 10, pp. 4270–4281, 2014.
- [10] Q. Wu, H. Li, F. Meng et al., "Blind image quality assessment based on multichannel feature fusion and label transfer," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 3, pp. 425–440, 2016.
- [11] X. Zhang, W. Lin, and P. Xue, "Just-noticeable difference estimation with pixels in images," *Journal of Visual Communication and Image Representation*, vol. 19, no. 1, pp. 30–41, 2008.
- [12] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion of the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no. 4, pp. 525–536, 1974.
- [13] V. Laparra, J. Muñoz-Marí, and J. Malo, "Divisive normalization image quality metric revisited," *Journal of the Optical Society of America A: Optics and Image Science, and Vision*, vol. 27, no. 4, pp. 852–864, 2010.
- [14] H. R. Sheikh and A. C. Bovik, "Information Theoretic Approaches to Image Quality Assessment," *Handbook of Image and Video Processing*, pp. 975–989, 2005.
- [15] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2117–2128, 2005.
- [16] D. V. Rao and L. P. Reddy, "Contrast weighted perceptual structural similarity index for image quality assessment," in *Proceedings of the IEEE India Council Conference, INDICON 2009*, India, December 2009.
- [17] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 5, pp. 1185–1198, 2011.
- [18] Y. Ding, Y. Zhang, X. Wang, X. Yan, and A. S. Krylov, "Perceptual image quality assessment metric using mutual information of Gabor features," *Science China Information Sciences*, vol. 57, no. 3, pp. 1–9, 2014.
- [19] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 21, no. 7, pp. 971–982, 2011.
- [20] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [21] M. P. Sampat, Z. Wang, S. Gupta, A. C. Bovik, and M. K. Markey, "Complex wavelet structural similarity: a new image similarity index," *IEEE Transactions on Image Processing*, vol. 18, no. 11, pp. 2385–2401, 2009.
- [22] C. Li and A. C. Bovik, "Content-partitioned structural similarity index for image quality assessment," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 517–526, 2010.
- [23] L. Zhang, L. Zhang, X. Mou, and D. Zhang, "FSIM: a feature similarity index for image quality assessment," *IEEE Transactions on Image Processing*, vol. 20, no. 8, pp. 2378–2386, 2011.
- [24] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.
- [25] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: a highly efficient perceptual image quality index," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 684–695, 2014.
- [26] J. Li, L. Zou, J. Yan, D. Deng, T. Qu, and G. Xie, "No-reference image quality assessment using Prewitt magnitude based on convolutional neural networks," *Signal, Image and Video Processing*, vol. 10, no. 4, pp. 609–616, 2016.
- [27] N. Kruger, P. Janssen, S. Kalkan et al., "Deep hierarchies in the primate visual cortex: what can we learn for computer vision?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1847–1871, 2013.
- [28] A. A. Nasr and M. Z. Abdulkageed, "An Efficient Reverse Engineering Hardware Trojan Detector Using Histogram of Oriented Gradients," *Journal of Electronic Testing*, vol. 33, no. 1, pp. 93–105, 2017.
- [29] H. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, *Live Image Quality Assessment Database Release 2*, 2014.
- [30] Z. Zhang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proceedings of the 37th Asilomar Conf. Signals Syst*, pp. 1398–1402, 2003.

- [31] D. M. Chandler, "Seven challenges in image quality assessment: past, present, and future research," *ISRN Signal Processing*, vol. 2013, Article ID 905685, 54 pages, 2013.
- [32] C. Chang and C. Lin, "LIBSVM: A Library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.
- [33] B. Scholkopf and A. Smola, *Learning with kernels*, MIT Press, Cambridge, MA, 2002.
- [34] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, and F. Battisti, "TID2008-A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics," *Advances of Modern Radioelectronics*, vol. 10, no. 4, pp. 30–45, 2009.
- [35] N. Ponomarenko, L. Jin, O. Ieremeiev et al., "Image database TID2013: peculiarities, results and perspectives," *Signal Processing: Image Communication*, vol. 30, pp. 57–77, 2015.
- [36] E. C. Larson and D. M. Chandler, *Categorical Image Quality (CSIQ) Database*, 2014, <http://vision.okstate.edu/csiq>.
- [37] M. Oszust, "Decision fusion for image quality assessment using an optimization approach," *IEEE Signal Processing Letters*, vol. 23, no. 1, pp. 65–69, 2016.
- [38] L. Li, H. Cai, Y. Zhang, W. Lin, A. C. Kot, and X. Sun, "Sparse representation-based image quality index with adaptive sub-dictionaries," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3775–3786, 2016.
- [39] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, Article ID 011006, 2010.
- [40] D. Temel and G. AlRegib, "CSV: Image quality assessment based on color, structure, and visual system," *Signal Processing: Image Communication*, vol. 48, pp. 92–103, 2016.
- [41] F. Wang, X. Sun, Z. Guo, Y. Huang, and K. Fu, "An Object-Distortion Based Image Quality Similarity," *IEEE Signal Processing Letters*, vol. 22, no. 10, pp. 1534–1537, 2015.

